Local Asymptotics of P-Spline Smoothing

Xiao Wang,* Jinglai Shen,† and David Ruppert‡

December 9, 2009

Abstract

This paper addresses asymptotic properties of general penalized spline estimators with an arbitrary B-spline degree and an arbitrary order difference penalty. The estimator is approximated by a solution of a linear differential equation subject to suitable boundary conditions. It is shown that, in certain sense, the penalized smoothing corresponds approximately to smoothing by the kernel method. The equivalent kernels for both inner points and boundary points are obtained with the help of Green's functions of the differential equation. Further, the asymptotic normality is established for the estimator at interior points. It is shown that the convergence rate is independent of the degree of the splines, and the number of knots does not affect the asymptotic distribution, provided that it tends to infinity fast enough.

Key Words: Difference penalty, equivalent kernel, Green's function, P-spline.

1 Introduction

Consider the problem of estimating the function $f : [0,1] \to \mathbb{R}$ from a univariate regression model $y_i = f(t_i) + \epsilon_i$, $i = 1, \dots, n$, where the t_i are pre-specified design points and the ϵ_i are iid normal random variables with mean 0 and variance σ^2. This paper presents a local asymptotic theory of penalized spline estimators of f.

The penalized spline regression model with difference penalty was introduced by Eilers and Marx (1996), who coined the term "P-splines", but using less knots for the regression problem can be traced back at least to O'Sullivan (1986). Penalized spline smoothing has become popular over the last decade and the uses of low rank bases lead to highly tractable computation. The methodology and applications of P-splines are discussed extensively in Ruppert, Wand and Carroll (2003). On the other hand, asymptotic properties of the P-spline estimators are less explored in the literature. A few exceptions include recent papers such as Hall and Opsomer (2005), Li and Ruppert (2008), and Claeskens, Krivobokova, and Opsomer (2009). Hall and Opsomer (2005) placed knots continuously over a design set and established consistency of the estimator. Li and Ruppert (2008) developed an asymptotic theory of P-splines for piecewise constant and linear B-splines with the first and second order

*Department of Statistics, Purdue University, West Lafayette, IN 47909, U.S.A. Email: wangxiao@purdue.edu.

†Department of Mathematics and Statistics, University of Maryland Baltimore County, Baltimore, MD 21250, U.S.A. Email: shenj@umbc.edu.

‡School of Operations Research and Information Engineering and Department of Statistical Sciences, Cornell University, Ithaca, NY 14853, U.S.A. Email: dr24@cornell.edu.

difference penalties. Claeskens, Krivobokova, and Opsomer (2009) studied bias, variance and asymptotic rates of the P-spline estimator under different choices of the number of knots and penalty parameters. An interested reader may also refer to Pal and Woodroofe (2007), Shen and Wang (2009), and Wang and Shen (2009) for shape constrained regression estimators and their applications.

The P-spline model approximates the regression function by $f^{[p]}(x) = \sum_{k=1}^{K_n+p} b_k B_k^{[p]}(x)$, where $\{B_k^{[p]} : k = 1, \ldots, K_n + p\}$ is the pth degree B-spline basis with knots $0 = \kappa_0 < \kappa_1 < \cdots < \kappa_{K_n} = 1$. The value of K_n will depend upon n as discussed below. The spline coefficients $\hat{b} = \{\hat{b}_k, k = 1, \ldots, K_n + p\}$ subject to the mth-order difference penalty are chosen to minimize

$$\sum_{i=1}^{n} \left[y_i - \sum_{k=1}^{K_n+p} b_k B_k^{[p]}(t_i)\right]^2 + \lambda^* \sum_{k=m+1}^{K_n+p} \left[\Delta^m(b_k)\right]^2, \tag{1}$$

where $\lambda^* > 0$ and Δ is the backward difference operator, i.e., $\Delta b_k \equiv b_k - b_{k-1}$ and

$$\Delta^m b_k = \Delta\Delta^{m-1} b_k = \cdots = \sum_{j=0}^{m} (-1)^{m-j} \binom{m}{j} b_{k-m+j}. \tag{2}$$

For simplicity, we assume that both the design points and the knots are equally spaced on the interval $[0, 1]$. We also assume that n/K_n is an integer denoted by M_n. Hence every M_nth design point is a knot, that is, $\kappa_j = t_{jM_n}$ for $j = 1, \ldots, K_n$; a more general case is discussed briefly in Section 6. The P-spline estimator is given by $\hat{f}^{[p]}(x) = \sum_{k=1}^{K_n+p} \hat{b}_k B_k^{[p]}(x)$.

This paper develops a general asymptotic theory of P-splines under an arbitrary choice of p and m. It is shown that the P-spline estimator can be approximated by the solution of an ordinary differential equation (ODE) with suitable boundary conditions. This estimator is then shown to be described by a kernel estimator, using a Green's function obtained from a closely related boundary value problem as a kernel. The asymptotic properties of the estimator thus are explicitly established based on the Green's function and the solution of the differential equation. It is worth mentioning that asymptotic analysis of smoothing splines using Green's functions was performed by Rice and Rosenblatt (1983), Silverman (1984), Messer (1991), Nychka (1995) and Pal and Woodroofe (2007). However, these papers only treat limited special cases. In contrast, the current paper develops a general framework for P-splines. This framework leads to a relatively simpler approach to obtain a closed-form expression of an equivalent kernel for both inner points and boundary points at the first time. Further, we show that the convergence rate of $\hat{f}^{[p]}$ depends only on m but not on p, as long as K_n tends to infinity fast enough; see Corollary 4.1 where K_n is of order n^γ, where $\gamma > (2m-1)/(4m+1)$.

The contributions of the present paper are twofold: (i) the paper develops a general approach for asymptotic analysis of a P-spline estimator with an arbitrary spline degree and arbitrary order difference penalty via Green's functions. To handle a general P-spline estimator, various techniques for linear ODEs are exploited to obtain a corresponding Green's function. (ii) the closed-form expressions of equivalent kernels for both inner and boundary points are established and convergence rates are developed for general P-spline estimators. Compared with the existing results based on matrix techniques, e.g. Li and Ruppert (2008) and Claeskens, Krivobokova, and Opsomer (2009), the use of Green's functions considerably simplifies the development and yields an instrumental alternative to establish the equivalent kernels for general P-splines. Moreover, this also leads to the convergence rates and the

observation that the rates are independent of the splines' degrees and the number of knots for an arbitrary P-spline estimator. While this observation is pointed out by Li and Ruppert (2008) for piecewise constant and piecewise linear splines and is conjectured for general P-splines, no rigorous justification has been given for general P-splines in the literature; the current paper offers a satisfactory answer to this issue in a general setting.

The paper is organized as follows. Section 2 characterizes the general P-spline estimator as an approximate solution of a linear differential equation subject to suitable boundary conditions. Section 3 investigates the solution of such the differential equation and obtains the related Green's functions as equivalent kernels for a P-spline estimator of an arbitrary B-spline degree with any order difference penalty. Using these Green's functions, the asymptotic properties of P-splines are established in Section 4. Section 5 addresses kernel approximation near the boundary of the design set. By formulating boundary conditions as an appropriate integral form, an explicit equivalent kernel is obtained. Finally, extensions to unequally spaced data and multivariate P-splines are discussed in Section 6.

2 Characterization of the estimator

Let $X = [B_k(x_i)] \in \mathbb{R}^{n\times(K_n+p)}$ be the design matrix, and let $D_m \in \mathbb{R}^{(K+p-m)\times(K+p)}$ be the mth-order difference matrix such that $D_m b = [\Delta^m(b_{m+1}), \ldots, \Delta^m(b_{K_n+p})]^T$. The optimality condition is given by

$$(X^T X + \lambda^* D_m^T D_m)\hat{b} = X^T y, \tag{3}$$

where $y = (y_1, \ldots, y_n)^T$.

To characterize the P-spline estimator $\hat{f}^{[p]}$, we introduce more notation. Define $C \in \mathbb{R}^{(K_n+p)\times(K_n+p)}$ and $\tilde{C} \in \mathbb{R}^{(K_n+p)\times n}$, respectively, as

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 1 & 0 & \cdots & 0 & 0 \\ & \cdots & & & \cdots & & \\ 1 & 1 & 1 & 1 & \cdots & 1 & 0 \\ 1 & 1 & 1 & 1 & \cdots & 1 & 1 \end{bmatrix} \quad \text{and} \quad \tilde{C} = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ & \cdots & & \cdots & & \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ \mathbf{1}^T & 0 & 0 & \cdots & 0 & 0 \\ \mathbf{1}^T & \mathbf{1}^T & 0 & \cdots & 0 & 0 \\ & \cdots & & \cdots & & \\ \mathbf{1}^T & \mathbf{1}^T & \mathbf{1}^T & \cdots & \mathbf{1}^T & 0 \\ \mathbf{1}^T & \mathbf{1}^T & \mathbf{1}^T & \cdots & \mathbf{1}^T & \mathbf{1}^T \end{bmatrix},$$

where $\mathbf{1} = [1, 1, \cdots, 1]^T \in \mathbb{R}^{M_n\times 1}$. Since C is invertible, for any $k \in \mathbb{N}$, (3) is equivalent to

$$\lambda^* C^k D_m^T D_m \hat{b} + C^k X^T \hat{f} = C^k X^T y, \tag{4}$$

where $\hat{f} = [\hat{f}^{[p]}(x_1), \ldots, \hat{f}^{[p]}(x_n)]^T$ and $C^k = \underbrace{CC\cdots C}_{k-\text{copies}}$. The matrix $D_m^T D_m$ is a banded symmetric matrix. Except for the first m and last m rows, every row of $D_m^T D_m$ has the form $(0, \cdots, 0, \omega_0^*, \omega_1^*, \cdots, \omega_{2m}^*, 0, \cdots, 0)$, where $\omega_j^* = (-1)^m(-1)^{2m-j}\binom{2m}{j}$, $j = 0, \ldots, 2m$. Moreover, except for the first $m-k$ and last m rows, the ith row of $C^k D_m^T D_m$ has the form

$$\Big(\quad \underbrace{0, \cdots, 0,}_{(i-m+k-1)-\text{copies}} \quad \omega_0, \cdots, \omega_{2m-k}, \quad \underbrace{0, \cdots, 0}_{(K_n+p)-(i+m)-\text{copies}} \quad \Big),$$

where

$$\omega_j = (-1)^m (-1)^{2m-k-j} \binom{2m-k}{j}, \; j = 0, \ldots, 2m-k. \tag{5}$$

Further, the elements of the last k rows of $C^k D_m^T D_m$ are all zeros. In particular, when $k = m$,

$$C^m D_m^T D_m \hat{b} = (-1)^m \left[\; \Delta^m \hat{b}_{m+1}, \Delta^m \hat{b}_{m+2}, \cdots, \Delta^m \hat{b}_{K_n+p}, 0, \ldots, 0 \; \right]^T. \tag{6}$$

It is also interesting to note the derivative formula for B-spline functions (de Boor, 2001)

$$\frac{d^l}{dx^l} \sum_{k=1}^{K_n+p} b_k B_k^{[p]}(x) = \sum_{k=l+1}^{K_n+p} K_n^l \Delta^l b_k \; B_{k-1}^{[p-l]}(x), \quad l \leq p. \tag{7}$$

Hence,

$$\frac{d^m}{dx^m} \sum_{k=1}^{K_n+m} \hat{b}_k B_k^{[m]}(x) = K_n^m \sum_{k=m+1}^{K_n+m} \Delta^m \hat{b}_k B_{k-m}^{[0]}(x),$$

and therefore,

$$\Delta^m \hat{b}_{m+k} = \frac{1}{K_n^m} \frac{d^m}{dx^m} \hat{f}^{[m]}(x), \qquad x \in (\kappa_{k-1}, \kappa_k], \quad k = 1, \ldots, K_n. \tag{8}$$

Let ω_1 be the uniform distribution on $x_1, \ldots, x_n$ and ω_2 be the uniform distribution on $\kappa_1, \ldots, \kappa_{K_n}$. Let g and $\check{f}$ be two piecewise constant functions for which $g(x_k) = y_k$ and $\check{f}(x_k) = \hat{f}(x_k)$ for $k = 1, ..., n$, respectively. Let $G_1(x) = \int_0^x g(t) d\omega_1(t)$, $\check{F}_1(x) = \int_0^x \check{f}(t) d\omega_1(t)$, $\hat{F}_1(x) = \int_0^x \hat{f}(t) dt$, and for $k \geq 2$, define

$$G_k(x) = \int_0^x G_{k-1}(t) d\omega_2(t), \qquad \check{F}_k(x) = \int_0^x \check{F}_{k-1}(t) d\omega_2(t), \qquad \hat{F}_k(x) = \int_0^x \hat{F}_{k-1}(t) dt.$$

To obtain the analogous representation for $\hat{f}$, we introduce a few variables and functions related to the true regression function f. Define $\Phi_1(x) = \int_0^x f(t) dt$, $\tilde{\Phi}_1(x) = \int_0^x f(t) d\omega_1(t)$, and for $k \geq 2$,

$$\Phi_k(x) = \int_0^x \Phi_{k-1}(t) dt, \qquad \tilde{\Phi}_k(x) = \int_0^x \tilde{\Phi}_{k-1}(t) d\omega_2(t).$$

Letting $R = C \; X^T - \tilde{C}$, we have $C^m X^T \hat{f} = C^{m-1} \tilde{C} \hat{f} + C^{m-1} R \hat{f}$. Therefore, the jth row of (4), when $k = m$, can be written as

$$\check{F}_m(\kappa_{j+p-1}) + R_{fj} + (-1)^m \frac{\lambda^*}{n K_n^{m-1}} \Delta^m b_{m+j} = G_m(\kappa_{j+p-1}) + R_{yj}, \quad j = 1, \ldots, K_n, \tag{9}$$

where R_{fj} and R_{yj} are the jth row of $\frac{1}{nK_n^{m-1}} C^{m-1} R \hat{f}$ and $\frac{1}{nK_n^{m-1}} C^{m-1} R y$, respectively. Furthermore, since the elements of the last k rows of $C^k D_m^T D_m$ are all zeros, we also have

$$\check{F}_k(1) = G_k(1), \qquad k = 1, \ldots, m. \tag{10}$$

Next, we proceed by replacing that difference equation (9) by an analogous differential equation. We shall focus on the case when $p = m$ first; the case when $p \neq m$ will be discussed in Section 4. For any $x \in [0, 1]$, letting $k_x = \lfloor K_n x \rfloor + 1$, (9) gives

$$\check{F}_m(\kappa_{k_x+p}) + R_{f,k_x+1} + (-1)^m \frac{\lambda^*}{n K_n^{m-1}} \Delta^m b_{m+k_x+1} = G_m(\kappa_{k_x+p}) + R_{y,k_x+1}. \tag{11}$$

Define

$$\tilde{R}(x) = \hat{F}_m(x) - G_m(x) + G_m(\kappa_{k_x+p}) - \check{F}_m(\kappa_{k_x+p}) + R_{y,k_x+1} - R_{f,k_x+1}. \tag{12}$$

Then, from (8) and (11), $\hat{F}_m$ solves the ordinary differential equation

$$(-1)^m \alpha \hat{F}_m^{(2m)}(x) + \hat{F}_m(x) = G_m(x) + \tilde{R}(x), \qquad 0 \le x \le 1, \tag{13}$$

where $\alpha = \lambda^*/(nK_n^{2m-1})$. We have $2m$ boundary conditions for (13):

$$\hat{F}_m^{(k)}(0) = 0, \quad \hat{F}_m^{(k)}(1) = G_{m-k}(1) + e_{m-k}, \quad k = 0, \dots, m-1,$$

where $e_{m-k} = \hat{F}_m^{(k)}(1) - \check{F}_{m-k}(1)$. We shall show that $\hat{f}^{[p]}$ is stochastically bounded, therefore the e_k are small with an order of $O_p(1/n)$.

3 Green's functions

The solution to (13) can be represented by a corresponding Green's function explicitly. It shall be shown that the P-spline estimator can be approximated by a kernel estimator, using the corresponding Green's function. For this end, consider the differential equation

$$(-1)^m \alpha F^{(2m)}(t) + F(t) = G(t), \qquad 0 \le t \le 1, \tag{14}$$

subject to the boundary conditions $F^{(i)}(0) = 0$ and $F^{(i)}(1) = G^{(i)}(1)$, $i = 0, \dots, m-1$. Let $\beta \equiv \alpha^{-1/(2m)}$. We consider two cases: (1) m is even; and (2) m is odd.

3.1 Even m

In this case, the characteristic equation is given by $\lambda^{2m} + \beta^{2m} = 0$, and we obtain $2m$ eigenvalues

$$\lambda_k = \beta\Big[\cos\frac{(1+2k)\pi}{2m} + \imath \sin\frac{(1+2k)\pi}{2m}\Big], \quad k = 0, 1, \cdots, 2m-1.$$

Let

$$\mu_k = \cos\frac{(1+2k)\pi}{2m} \quad \text{and} \quad \omega_k = \sin\frac{(1+2k)\pi}{2m}.$$

Then the homogeneous ODE: $\alpha F^{(2m)}(t) + F(t) = 0$ has $2m$ solutions

$$e^{(\pm\mu_k \pm \imath\omega_k)\beta t} = e^{\pm\beta\mu_k t}\left[\cos(\beta\omega_k t) \pm \imath \sin(\beta\omega_k t)\right], \quad k = 0, \cdots, \frac{m}{2} - 1,$$

where $\mu_k > 0$ and $\omega_k > 0$ for $k = 0, \cdots, \frac{m}{2} - 1$.

To find the corresponding Green's function for the ODE: $\alpha F^{(2m)}(t) + F(t) = G(t)$ on $[0, 1]$, we define the following function

$$L(t) \equiv \sum_{k=0}^{\frac{m}{2}-1} \beta\, e^{-\beta\mu_k t}\big[\, c_k \cos(\omega_k\beta t) + d_k \sin(\omega_k\beta t)\big], \tag{15}$$

where the coefficients c_k, d_k are to be determined, and $K(t,s) \equiv L(|t-s|)$. Since L is a linear combination of the solutions of the homogeneous ODE, $L^{(2m)} + \beta^{2m} L = 0$ also holds. Let

$$F_0(t) \equiv \int_0^1 K(t,s)G(s)ds, \quad t \in [0,1].$$

By noting $F_0(t) = \int_0^t L(t-s)G(s)ds + \int_t^1 L(s-t)G(s)ds$ for all $t \in [0,1]$, it is easy to verify that if

$$L^{(k)}(t)\big|_{t=0} = 0, \quad \forall \;\; k = 1,3,\cdots,2m-3, \quad \text{and} \quad L^{(2m-1)}(t)\big|_{t=0} = \frac{\beta^{2m}}{2}, \tag{16}$$

then $F_0(t)$ is a solution of $\alpha F^{(2m)} + F = G$.

To find the coefficients c_k, d_k, define

$$p_k(t) \equiv e^{-\beta\mu_k t}\big[c_k \cos(\omega_k\beta t) + d_k \sin(\omega_k\beta t)\big], \quad q_k(t) \equiv e^{-\beta\mu_k t}\big[-c_k \sin(\omega_k\beta t) + d_k \cos(\omega_k\beta t)\big].$$

Hence $p_k(0) = c_k$ and $q_k(0) = d_k$. Since

$$\begin{pmatrix} p'_k(t) \\ q'_k(t) \end{pmatrix} = \beta \underbrace{\begin{bmatrix} -\mu_k & \omega_k \\ -\omega_k & -\mu_k \end{bmatrix}}_{A_k} \begin{pmatrix} p_k(t) \\ q_k(t) \end{pmatrix}, \tag{17}$$

we have

$$\begin{pmatrix} p_k^{(j)}(t) \\ q_k^{(j)}(t) \end{pmatrix} = (\beta A_k)^j \begin{pmatrix} p_k(t) \\ q_k(t) \end{pmatrix},$$

where $p_k^{(j)}(t)$ and $q_k^{(j)}(t)$ stand for the j-th derivatives of p_k and q_k respectively. Letting $A_k^j(i,\ell)$ denote the (i,ℓ)-element of A_k^j, we obtain the following linear equation for $\{c_k, d_k\}$ from (16):

$$\underbrace{\begin{bmatrix} A_0(1,1) & A_0(1,2) & \cdots & \cdots & A_{\frac{m}{2}-1}(1,1) & A_{\frac{m}{2}-1}(1,2) \\ A_0^3(1,1) & A_0^3(1,2) & \cdots & \cdots & A^3_{\frac{m}{2}-1}(1,1) & A^3_{\frac{m}{2}-1}(1,2) \\ \vdots & \vdots & & & \vdots & \vdots \\ A_0^{(2m-3)}(1,1) & A_0^{(2m-3)}(1,2) & \cdots & \cdots & A^{(2m-3)}_{\frac{m}{2}-1}(1,1) & A^{(2m-3)}_{\frac{m}{2}-1}(1,2) \\ A_0^{(2m-1)}(1,1) & A_0^{(2m-1)}(1,2) & \cdots & \cdots & A^{(2m-1)}_{\frac{m}{2}-1}(1,1) & A^{(2m-1)}_{\frac{m}{2}-1}(1,2) \end{bmatrix}}_{A^e} \begin{bmatrix} c_0 \\ d_0 \\ \vdots \\ c_{\frac{m}{2}-1} \\ d_{\frac{m}{2}-1} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \frac{1}{2} \end{bmatrix}. \tag{18}$$

It shall be shown in Lemma 3.1 that the above equation has a unique solution.

3.2 Odd m

The characteristic equation is given by $\lambda^{2m} - \beta^{2m} = 0$ and the eigenvalues are:

$$\lambda_k = \beta\Big(\cos\frac{k\pi}{m} + \imath \sin\frac{k\pi}{m}\Big), \quad k = 0,1,\cdots,2m-1.$$

Then the homogeneous ODE: $\alpha F^{(2m)}(t) + F(t) = 0$ has $2m$ solutions: $e^{\pm\beta t}$ and

$$e^{(\pm\mu_k \pm \imath\omega_k)\beta t} = e^{\pm\beta\mu_k t}\left[\cos(\beta\omega_k t) \pm \imath \sin(\beta\omega_k t)\right], \quad k = 1,\cdots,\frac{m-1}{2},$$

where $\mu_k = \cos\frac{k\pi}{m} > 0$ and $\omega_k = \sin\frac{k\pi}{m} > 0$ for $k = 1, \cdots, \frac{m-1}{2}$. Similar to the even case, define

$$P(t) \equiv c_0\,\beta e^{-\beta t} + \sum_{k=1}^{(m-1)/2} \beta\, e^{-\beta\mu_k t}\big[\, c_k\, \cos(\omega_k\beta t) + d_k\, \sin(\omega_k\beta t)\big], \tag{19}$$

where the coefficients c_k, d_k are to be determined, and $P(t)$ satisfies $P^{(2m)}(t) - \beta^{2m}P(t) = 0$. Let $K(t,s) \equiv P(|t-s|)$ and $F_0(t) \equiv \int_0^1 K(t,s)G(s)ds$. It can be verified that if

$$P^{(k)}(t)\big|_{t=0} = 0, \quad \forall \;\; k = 1, 3, \cdots, 2m-3, \quad \text{and} \quad P^{(2m-1)}(t)\big|_{t=0} = -\frac{\beta^{2m}}{2}, \tag{20}$$

then $F_0(t)$ is a solution of $\alpha F^{(2m)} - F = -G$. Similarly, it can be shown that P is also a $2m$th-order kernel. To find the coefficients c_0 and c_k, d_k, we may use p_k, q_k and A_k introduced in the last subsection. Indeed, we obtain the following linear equation for c_0 and $\{c_k, d_k\}$ from (20):

$$\underbrace{\begin{bmatrix} -1 & A_1(1,1) & A_1(1,2) & \cdots & \cdots & A_{\frac{m-1}{2}}(1,1) & A_{\frac{m-1}{2}}(1,2) \\ -1 & A_1^3(1,1) & A_1^3(1,2) & \cdots & \cdots & A^3_{\frac{m-1}{2}}(1,1) & A^3_{\frac{m-1}{2}}(1,2) \\ \vdots & \vdots & \vdots & & & \vdots & \vdots \\ -1 & A_1^{(2m-3)}(1,1) & A_1^{(2m-3)}(1,2) & \cdots & \cdots & A^{(2m-3)}_{\frac{m-1}{2}}(1,1) & A^{(2m-3)}_{\frac{m-1}{2}}(1,2) \\ -1 & A_1^{(2m-1)}(1,1) & A_1^{(2m-1)}(1,2) & \cdots & \cdots & A^{(2m-1)}_{\frac{m-1}{2}}(1,1) & A^{(2m-1)}_{\frac{m-1}{2}}(1,2) \end{bmatrix}}_{A^o} \begin{bmatrix} c_0 \\ c_1 \\ d_1 \\ \vdots \\ c_{\frac{m-1}{2}} \\ d_{\frac{m-1}{2}} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ -\frac{1}{2} \end{bmatrix}. \tag{21}$$

3.3 The equivalent kernels

Lemma 3.1. *Each of the equations (18) and (21) has a unique solution.*

Proof. We introduce some trigonometric identities to be used in the proof. Let $p, q \in \mathbb{N}$. By observing $\sin(-\theta)\sum_{k=1}^{p}\cos[(2k-1)\theta] = \frac{1}{2}\sum_{k=1}^{p}\big[\sin(2(k-1)\theta) + \sin(-2k\theta)\big]$ and $\sin(\theta)\sum_{k=1}^{p}\sin[(2k-1)\theta] = \frac{1}{2}\sum_{k=1}^{p}\big[\cos(2(k-1)\theta) - \cos(2k\theta)\big]$, it is easy to see (i) for $\theta = \dfrac{q}{2p}\pi$, $\sum_{k=1}^{p}\cos[(2k-1)\theta] = 0$; and (ii) for $\theta = \dfrac{q}{p}\pi$, $\sum_{k=1}^{p}\sin[(2k-1)\theta] = 0$.

We consider an even m first. Let $\theta \equiv \pi - \dfrac{\pi}{2m}$. It is clear that $-\mu_k = \cos\big((2k+1)\theta\big)$ and $\omega_k = \sin\big((2k+1)\theta\big)$ for all $k = 0, \cdots, m/2-1$. Hence A_k in (17) becomes $A_k = M((2k+1)\theta)$, where $M(\cdot) \in \mathrm{SO}(2)$ is given by

$$M(\cdot) \equiv \begin{bmatrix} \cos(\cdot) & \sin(\cdot) \\ -\sin(\cdot) & \cos(\cdot) \end{bmatrix}. \tag{22}$$

Thus $(A_k)^j = M(j(2k+1)\theta)$. Let $A^e_{i\bullet}$ denote the ith row of A^e and $\eta_i \equiv (2i-1)\theta$. Hence,

$$A^e_{i\bullet} = \big(\; \cos(\eta_i) \quad \sin(\eta_i) \quad \cos(3\eta_i) \quad \sin(3\eta_i) \quad \cdots \quad \cos((m-1)\eta_i) \quad \sin((m-1)\eta_i) \;\big).$$

Therefore, $A^e_{i\bullet}\left(A^e_{i\bullet}\right)^T = \frac{m}{2}$, and if $i \neq j$, then

$$\begin{aligned}
A^e_{i\bullet}\left(A^e_{j\bullet}\right)^T &= \sum_{\ell=1}^{\frac{m}{2}} \Big[\cos((2\ell-1)(2i-1)\theta)\,\cos((2\ell-1)(2j-1)\theta) \\
&\qquad\qquad + \sin((2\ell-1)(2i-1)\theta)\,\sin((2\ell-1)(2j-1)\theta)\Big] \\
&= \sum_{\ell=1}^{\frac{m}{2}} \cos\big(2(2\ell-1)(i-j)\theta\big) \;=\; \sum_{\ell=1}^{\frac{m}{2}} \cos\big((2\ell-1)(i-j)\frac{\pi}{m}\big) \;=\; 0,
\end{aligned}$$

where the last step is attained from (i). This shows that $A^e(A^e)^T = \frac{m}{2}I$. Thus A^e is invertible so that equation (18) has a unique solution.

We then consider an odd m. In this case, $-\mu_k = \cos\left(\pi - \frac{k\pi}{m}\right)$ and $\omega_k = \sin\left(\pi - \frac{k\pi}{m}\right)$ for $k = 1, \cdots, (m-1)/2$. Let $\gamma_k \equiv \pi - \frac{k\pi}{m}$. Then the ith row of A^o is given by

$$\begin{aligned}
A^o_{i\bullet} \;=\; \Big(&\cos((2i-1)\pi) \quad \cos((2i-1)\gamma_1) \quad \sin((2i-1)\gamma_1) \quad \cos((2i-1)\gamma_2) \quad \sin((2i-1)\gamma_2) \\
&\cdots \quad \cdots \quad \cos\big((2i-1)\gamma_{\frac{m-1}{2}}\big) \quad \sin\big((2i-1)\gamma_{\frac{m-1}{2}}\big)\Big).
\end{aligned}$$

Let $A^o_{\bullet i}$ denote the ith column of A^o. Clearly $\left(A^o_{\bullet i}\right)^T A^o_{\bullet i} > 0$. For $i \neq j$, either $\left(A^o_{\bullet i}\right)^T A^o_{\bullet j} = \sum_{k=1}^m \cos((2k-1)\gamma_s)\cos((2k-1)\gamma_t)$ with $s \neq t$ or $\left(A^o_{\bullet i}\right)^T A^o_{\bullet j} = \sum_{k=1}^m \cos((2k-1)\gamma_s)\sin((2k-1)\gamma_t)$, for some $s, t \in \{1, \cdots, \frac{m-1}{2}\}$. Since

$$\begin{aligned}
\sum_{k=1}^m \cos\big((2k-1)\gamma_s\big)\cos\big((2k-1)\gamma_t\big) &= \frac{1}{2}\sum_{k=1}^m \Big[\cos((2k-1)(\gamma_s+\gamma_t)) + \cos((2k-1)(\gamma_s-\gamma_t))\Big], \\
\sum_{k=1}^m \cos\big((2k-1)\gamma_s\big)\sin\big((2k-1)\gamma_t\big) &= \frac{1}{2}\sum_{k=1}^m \Big[\sin((2k-1)(\gamma_s+\gamma_t)) + \sin((2k-1)(\gamma_s-\gamma_t))\Big],
\end{aligned}$$

we conclude that $\left(A^o_{\bullet i}\right)^T A^o_{\bullet j} = 0$ by using (i)–(ii) established at the beginning of the proof. This shows that $(A^o)^T A^o$ is a diagonal matrix with positive diagonal entries. Therefore A^o is invertible and equation (21) has a unique solution. $\square$

The following proposition show that L and P derived above yield the equivalent kernels.

Proposition 3.1. *When $\beta = 1$, $L(|t|)$ in (15) and $P(|t|)$ in (19) are $2m$th order kernels respectively.*

Proof. We consider $L(|t|)$ only since the other case follows from the similar argument. We shall show that $\int_{-\infty}^{\infty} L(|\tau|)d\tau = 1$ and $\int_{-\infty}^{\infty} \tau^k L(|\tau|)d\tau = 0$ for all $k = 1, \cdots, 2m-1$. This holds true trivially when k is odd. For an even k, by observing $L^{(2m)} + \beta^{2m}L = 0$ (with $\beta = 1$), we have

$$\int_{-\infty}^{\infty} \tau^k L(|\tau|)d\tau = 2\int_0^{\infty} \tau^k L(\tau)d\tau = -2\int_0^{\infty} \tau^k L^{(2m)}(\tau)d\tau.$$

Repeatedly using the integration by part, we deduce

$$\int_0^t \tau^k L^{(2m)}(\tau)d\tau = \sum_{i=0}^k \frac{k!}{(k-i)!}(-1)^i t^{k-i}\Big(L^{(2m-1-i)}(t) - L^{(2m-1-i)}(0)\Big).$$

In light of (16), we obtain the desired result. $\square$

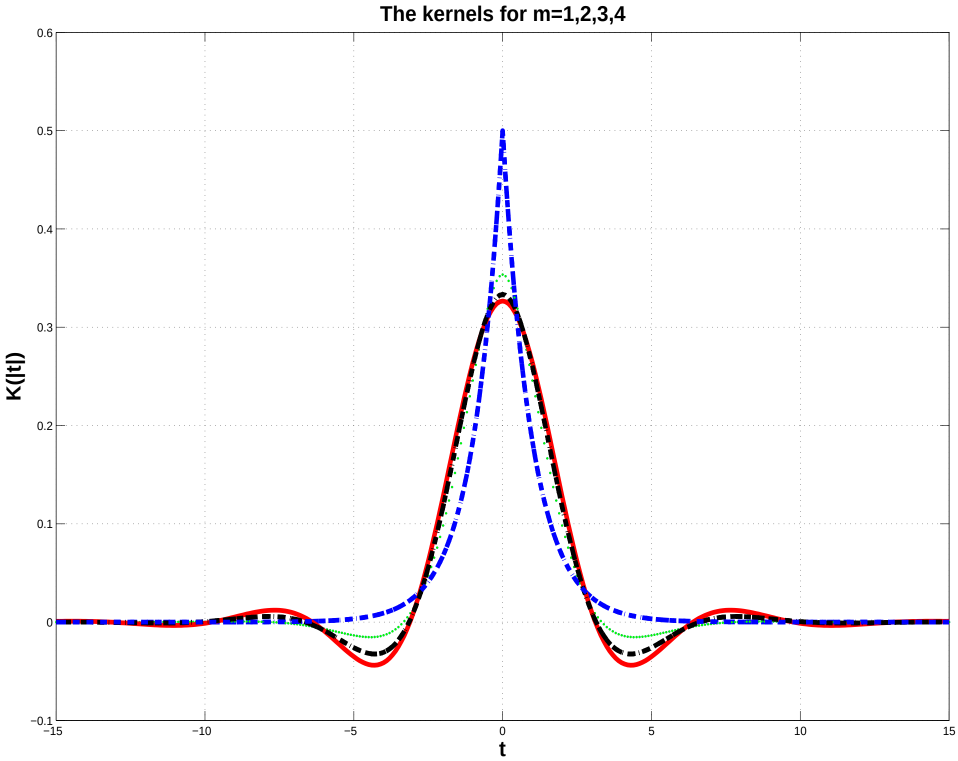

Figure 1: The equivalent kernels for $m = 1, 2, 3, 4$: (i) $m = 1$: the dashed line; (ii) $m = 2$: the dotted line; (iii) $m = 3$: the dashdot line; (iv) $m = 4$: the solid line.

Example 3.1. As an illustration, the closed-form expressions of the first four equivalent kernels are given below and their plots are shown in Figure 1, respectively.

$$
\begin{aligned}
m = 1: \quad K(t) &= \frac{1}{2} e^{-|t|} \\
m = 2: \quad K(t) &= \frac{1}{2\sqrt{2}} e^{-\frac{1}{\sqrt{2}}|t|} \Big(\cos \frac{|t|}{\sqrt{2}} + \sin \frac{|t|}{\sqrt{2}} \Big) \\
m = 3: \quad K(t) &= \frac{1}{6} e^{-|t|} + e^{-\frac{1}{2}|t|} \Big(\frac{1}{6} \cos \frac{\sqrt{3}|t|}{2} + \frac{\sqrt{3}}{6} \sin \frac{\sqrt{3}|t|}{2} \Big) \\
m = 4: \quad K(t) &= e^{-0.9239|t|} \Big(0.2310 \cos(0.3827|t|) + 0.0957 \sin(0.3827|t|) \Big) \\
&\quad + e^{-0.3827|t|} \Big(0.0957 \cos(0.9239|t|) + 0.2310 \sin(0.9239|t|) \Big)
\end{aligned}
$$

3.4 Boundary conditions

Recall that the boundary conditions for the ODE (14) are $F^{(i)}(0) = 0$, $F^{(i)}(1) = G^{(i)}(1)$, $i = 0, \cdots, m-1$. In the following, we consider an even m first. In this case, the homogeneous

ODE: $F^{(2m)} + \beta^{2m} F = 0$ has the following $2m$ (linearly independent) solutions:

$$e^{-\beta\mu_k t}\cos(\beta\omega_k t), \quad e^{-\beta\mu_k t}\sin(\beta\omega_k t), \quad e^{-\beta\mu_k (1-t)}\cos(\beta\omega_k t), \quad e^{-\beta\mu_k (1-t)}\sin(\beta\omega_k t),$$

where $k = 0, \cdots, \frac{m}{2} - 1$ and $\mu_k, \omega_k > 0$ for the above k. The solution to ODE (14) subject to the boundary conditions can be written as

$$F(t) = \underbrace{\int_0^1 L(|t-s|)G(s)ds}_{F_0(t)} + J(t), \tag{23}$$

where

$$J(t) = \sum_{k=0}^{\frac{m}{2}-1} \Big\{ e^{-\beta\mu_k t}\big[a_k \cos(\beta\omega_k t) + b_k \sin(\beta\omega_k t)\big] + e^{-\beta\mu_k(1-t)}\big[a_k^+ \cos(\beta\omega_k t) + b_k^+ \sin(\beta\omega_k t)\big] \Big\}, \tag{24}$$

and the coefficients a_k, b_k, a_k^+, b_k^+ are to be determined from the boundary conditions, and the kernel L is given in (15). Define $\|G\| \equiv \sup_{t\in[0,1]} |G(t)|$. Let $\mathbf{G} = \big(\|G\|, G(1), G'(1), \cdots, G^{(m-1)}(1)\big)$, and

$$\mathbf{a} = \Big(a_0, b_0, \cdots, a_{\frac{m}{2}-1}, b_{\frac{m}{2}-1}, a_0^+, b_0^+, \cdots, a^+_{\frac{m}{2}-1}, b^+_{\frac{m}{2}-1} \Big)^T \tag{25}$$

be the coefficient vector.

By making use of the boundary conditions, we obtain the linear equation $B^e\mathbf{a} = \mathbf{v}$, where $\mathbf{v}^T = [\mathbf{v_0}, \mathbf{v_1}]$,

$$\mathbf{v_0} = \Big[-F_0(0),\ -\frac{F_0'(0)}{\beta}, \cdots, -\frac{F_0^{(m-1)}(0)}{\beta^{m-1}} \Big],$$

$$\mathbf{v_1} = \Big[-F_0(1) + G(1),\ \frac{-F_0'(1) + G'(1)}{\beta}, \cdots, \frac{-F_0^{(m-1)}(1) + G^{(m-1)}(1)}{\beta^{m-1}} \Big],$$

and

$$B^e = \begin{bmatrix} B^e_{11} & B^e_{12} \\ B^e_{21} & B^e_{22} \end{bmatrix}.$$

Here the matrix blocks $B^e_{ij} \in \mathbb{R}^{m\times m}$ are obtained via the similar technique in Section 3.1 as

$$B^e_{11} = \begin{bmatrix} 1 & 0 & 1 & 0 & \cdots & \cdots & 1 & 0 \\ \cos(\eta_1) & \sin(\eta_1) & \cos(3\eta_1) & \sin(3\eta_1) & \cdots & \cdots & \cos((m-1)\eta_1) & \sin((m-1)\eta_1) \\ \vdots & \vdots & \vdots & \vdots & & & \vdots & \vdots \\ \cos(\eta_{m-1}) & \sin(\eta_{m-1}) & \cos(3\eta_{m-1}) & \sin(3\eta_{m-1}) & \cdots & \cdots & \cos((m-1)\eta_{m-1}) & \sin((m-1)\eta_{m-1}) \end{bmatrix}, \tag{26}$$

where $\eta_k = k(\pi - \frac{\pi}{2m})$, $k = 1, \cdots, m-1$, and

$$B^e_{22} = \begin{bmatrix} \cos(\psi_{0,0}) & \sin(\psi_{0,0}) & \cos(\psi_{1,0}) & \sin(\psi_{1,0}) & \cdots & \cdots & \cos(\psi_{\frac{m}{2}-1,0}) & \sin(\psi_{\frac{m}{2}-1,0}) \\ \cos(\psi_{0,1}) & \sin(\psi_{0,1}) & \cos(\psi_{1,1}) & \sin(\psi_{1,1}) & \cdots & \cdots & \cos(\psi_{\frac{m}{2}-1,1}) & \sin(\psi_{\frac{m}{2}-1,1}) \\ \vdots & \vdots & \vdots & \vdots & & & \vdots & \vdots \\ \cos(\psi_{0,m-1}) & \sin(\psi_{0,m-1}) & \cos(\psi_{1,m-1}) & \sin(\psi_{1,m-1}) & \cdots & \cdots & \cos(\psi_{\frac{m}{2}-1,m-1}) & \sin(\psi_{\frac{m}{2}-1,m-1}) \end{bmatrix},$$

where $\eta_k^+ = \frac{k\pi}{2m}$ for $k = 0, \cdots, m-1$, and $\psi_{j,\ell} = \beta\,\omega_j + (2j+1)\eta_\ell^+$ for all $j = 0, \cdots, \frac{m}{2}$, $\ell = 0, \cdots, m-1$, and each entry of B_{12}^e and B_{21}^e is of order $O(e^{-\beta})$.

Lemma 3.2. *Given an even m. There exist positive real numbers β_* and ϱ, dependent on m only, such that for all $\beta \geq \beta_*$, the coefficient vector $\mathbf{a}$ is unique and satisfies $\|\mathbf{a}\| \leq \varrho\,\|\mathbf{G}\|$.*

Proof. Note that for β sufficiently large, each element of B_{12}^e and B_{21}^e is sufficiently small. Hence it suffices to show that B_{11}^e and B_{22}^e are invertible. For this end, let $B_{11}^e(k)$ denote the kth column of B_{11}^e. Define $C_{11} \equiv \big[B_{11}^e(2)\ B_{11}^e(1)\ B_{11}^e(4)\ B_{11}^e(3)\ \cdots\ B_{11}^e(m)\ B_{11}^e(m-1)\big]$. Letting $\vartheta = \frac{\pi}{2m}$, it can be verified that

$$B_{11}^e + \imath C_{11} = \begin{bmatrix} 1 & \imath & 1 & \imath & \cdots & \cdots & 1 & \imath \\ -e^{-\imath\vartheta} & -\imath e^{\imath\vartheta} & -e^{-\imath 3\vartheta} & -\imath e^{\imath 3\vartheta} & \cdots & \cdots & -e^{-\imath(m-1)\vartheta} & -\imath e^{\imath(m-1)\vartheta} \\ \vdots & \vdots & \vdots & \vdots & & & \vdots & \vdots \\ (-e^{-\imath\vartheta})^{m-1} & \imath(-e^{\imath\vartheta})^{m-1} & (-e^{-\imath 3\vartheta})^{m-1} & \imath(-e^{\imath 3\vartheta})^{m-1} & \cdots & \cdots & (-e^{-\imath(m-1)\vartheta})^{m-1} & \imath(-e^{\imath(m-1)\vartheta})^{m-1} \end{bmatrix}.$$

Therefore $B_{11}^e + \imath C_{11}$ can be written as $\mathrm{diag}(1, \imath, 1, \imath, \cdots, 1, \imath)V$, where V is an invertible Vandermonde matrix. This implies that $B_{11}^e + \imath C_{11}$ is invertible. On the other hand, by noting $C_{11} = B_{11}^e J$, where $J = \mathrm{diag}\,\underbrace{(J_*, \cdots, J_*)}_{\frac{m}{2}-\text{copies}}$ with $J_* = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, $B_{11}^e + \imath C_{11} = B_{11}^e(I + \imath J)$. It is easily seen that $I + \imath J$ is invertible, so is B_{11}^e. To show the invertibility of B_{22}^e, it is noticed that $B_{22}^e = \widetilde{B}_{22}^e R$, where $\widetilde{B}_{22}^e$ is similar to B_{11}^e defined in (26) with η_i replaced by η_i^+ and $R = \mathrm{diag}\big(M(\beta\omega_0), M(\beta\omega_1), \cdots, M(\beta\omega_{\frac{m}{2}})\big)$ where $M(\cdot)$ is given in (22). Clearly R is invertible for all β, and it can be proved in the similar way as for B_{11}^e that $\widetilde{B}_{22}^e$ is nonsingular. Hence, B_{22}^e is invertible for all β. Consequently $\det(B^e) = \det(B_{11}^e)\det(B_{22}^e) + O(e^{-m\beta}) \neq 0$ for all β sufficiently large. In addition, since each entry of the adjoint of B^e is bounded, we deduce that $(B^e)^{-1} = \frac{\mathrm{adj}(B^e)}{\det(B^e)}$ is bounded and the upper bound depends on m only, where $\mathrm{adj}(B^e)$ stands for the adjoint of B^e. Furthermore, letting $\kappa = \max_k(|c_k|, |d_k|)$, where c_k, d_k are the coefficients in the kernel L, and $\varrho = \min\{\mu_k,\ k = 0, \cdots, \frac{m}{2} - 1\}$, we have, for $t_* = 0$ or 1,

$$\frac{\big|F_0^{(j)}(t_*)\big|}{\beta^j} \leq 2m\kappa \int_0^1 \beta e^{-\beta\varrho\tau} d\tau\, \|G\| \leq 2m\kappa/\varrho\, \|G\|, \qquad \forall\ j = 1, \cdots, m-1.$$

As a result, the equation $B^e x = \mathbf{v}$ has a unique solution $\mathbf{a}$ that satisfies the desired bound. $\square$

Consider an odd m. The homogeneous ODE: $F^{(2m)} - \beta^{2m}F = 0$ has the following $2m$ (linearly independent) solutions:

$$e^{\pm\beta t}, \quad e^{-\beta\mu_k t}\cos(\beta\omega_k t), \quad e^{-\beta\mu_k t}\sin(\beta\omega_k t), \quad e^{-\beta\mu_k(1-t)}\cos(\beta\omega_k t), \quad e^{-\beta\mu_k(1-t)}\sin(\beta\omega_k t),$$

where $k = 1, \cdots, \frac{m-1}{2}$ and $\mu_k, \omega_k > 0$ for the above k. The solution to ODE (14) subject to the boundary conditions can be written as

$$F(t) = \underbrace{\int_0^1 P(|t-s|)G(s)ds}_{F_0(t)} + J(t), \tag{27}$$

where

$$J(t) = a_0 e^{-\beta t} + a_0^+ e^{-\beta(1-t)} + \sum_{k=1}^{\frac{m-1}{2}} \Big\{ e^{-\beta\mu_k t}\big[a_k \cos(\beta\omega_k t) + b_k \sin(\beta\omega_k t)\big] + e^{-\beta\mu_k(1-t)}\big[a_k^+ \cos(\beta\omega_k t) + b_k^+ \sin(\beta\omega_k t)\big]\Big\}, \tag{28}$$

and the coefficients a_k, b_k, a_k^+, b_k^+ are to be determined from the boundary conditions, and the kernel P is given in (19). Let

$$\mathbf{b} = \Big(a_0, a_1, b_1, \cdots, a_{\frac{m-1}{2}}, b_{\frac{m-1}{2}}, a_0^+, a_1^+, b_1^+, \cdots, a^+_{\frac{m-1}{2}}, b^+_{\frac{m-1}{2}}\Big)^T \tag{29}$$

be the coefficient vector. Similar to the case where m is even, we obtain the linear equation $B^o\mathbf{b} = \mathbf{v}$, where $\mathbf{v}^T = [\mathbf{v_0}, \mathbf{v_1}]$ and

$$B^o = \begin{bmatrix} B^o_{11} & B^o_{12} \\ B^o_{21} & B^o_{22} \end{bmatrix}.$$

Here the matrix blocks $B^o_{ij} \in \mathbb{R}^{m\times m}$ are obtained via the similar technique in Section 3.2 as

$$B^o_{11} = \begin{bmatrix} 1 & 1 & 0 & \cdots & \cdots & 1 & 0 \\ -1 & \cos(\gamma_1) & \sin(\gamma_1) & \cdots & \cdots & \cos(\gamma_{\frac{m-1}{2}}) & \sin(\gamma_{\frac{m-1}{2}}) \\ \vdots & \vdots & \vdots & & & \vdots & \vdots \\ (-1)^{m-1} & \cos((m-1)\gamma_1) & \sin((m-1)\gamma_1) & \cdots & \cdots & \cos((m-1)\gamma_{\frac{m-1}{2}}) & \sin((m-1)\gamma_{\frac{m-1}{2}}) \end{bmatrix}, \tag{30}$$

where $\gamma_k = (\pi - \frac{k\pi}{m}), k = 1, 2, \cdots, \frac{m-1}{2}$, and

$$B^o_{22} = \begin{bmatrix} 1 & \cos(\zeta_{1,0}) & \sin(\zeta_{1,0}) & \cdots & \cdots & \cos(\zeta_{\frac{m-1}{2},0}) & \sin(\zeta_{\frac{m-1}{2},0}) \\ 1 & \cos(\zeta_{1,1}) & \sin(\zeta_{1,1}) & \cdots & \cdots & \cos(\zeta_{\frac{m-1}{2},1}) & \sin(\zeta_{\frac{m-1}{2},1}) \\ \vdots & \vdots & \vdots & & & \vdots & \vdots \\ 1 & \cos(\zeta_{1,m-1}) & \sin(\zeta_{1,m-1}) & \cdots & \cdots & \cos(\zeta_{\frac{m-1}{2},m-1}) & \sin(\zeta_{\frac{m-1}{2},m-1}) \end{bmatrix},$$

where $\gamma_k^+ = \frac{k\pi}{m}$, $\zeta_{k,\ell} = \beta\,\omega_k + \ell\gamma_k^+$ for all $k = 1, 2, \cdots, \frac{m-1}{2}$, $\ell = 0, 1, \cdots, m-1$, and each entry of B^o_{12} and B^o_{21} is of order $O(e^{-\beta})$. To show the invertibility of B^o_{11}, we introduce $E_{11} \equiv \big[0\ B^o_{11}(3)\ B^o_{11}(2)\ B^o_{11}(5)\ B^o_{11}(4)\ \cdots\ B^o_{11}(m)\ B^e_{11}(m-1)\big]$, where $B^o_{11}(k)$ denotes the kth column of B^o_{11}. As before it can be shown that $B^o_{11} + \imath E_{11}$ is nonsingular and $B^o_{11} + \imath E_{11} = B^o_{11}(I + \imath K)$, where $K \equiv \mathrm{diag}(0, \underbrace{J_*, \cdots, J_*}_{\frac{m-1}{2}-\text{copies}})$ and J_* is the 2×2 matrix defined before. Since $I + \imath K$ is nonsingular, so is B^o_{11}. Furthermore, by applying the similar technique, we can show that B^o_{22} is invertible for all β. This thus implies that for all β sufficiently large, B^o is invertible and each entry of $(B^o)^{-1}$ is bounded by a positive number depending on m only. We summarize the above discussions as follows:

Lemma 3.3. *Given an odd m. There exist positive real numbers β_* and ϱ, dependent on m only, such that for all $\beta \geq \beta_*$, the coefficient vector $\mathbf{b}$ is unique and satisfies $\|\mathbf{b}\| \leq \varrho\,\|\mathbf{G}\|$.*

4 Asymptotic properties of P-splines

To establish the asymptotic properties of the estimator, we first represent $\hat{F}_m$ as the sum of the convolutions of $K(t,s)$ (defined in Section 3) with G_m and a remainder term that is of smaller order.

Lemma 4.1. *The $\hat{F}_m$ in (13) can be represented as*

$$\hat{F}_m(t) = \int_0^1 K(s,t)G_m(s)ds + \int_0^1 K(s,t)\tilde{R}(s)ds + J(t), \tag{31}$$

where $J(t)$ is given by (24) and (28) for even m and odd m, respectively. The $\|\cdot\|_\infty$-norms of both coefficient vectors $\mathbf{a}$ in (25) and $\mathbf{b}$ in (29) are stochastically bounded, and $\|\tilde{R}\| = O_p\Big((\frac{\log K_n}{nK_n})^{1/2}\Big)$.

Proof. The representation of $\hat{F}_m$ in (31) follows from the discussions in Section 3. The stochastic boundedness of the coefficient vectors is the direct applications of Lemma 3.2 and Lemma 3.3. Let $\bar{y} = \frac{K_n}{n}X^Ty$ and $\lambda = \lambda^* K_n/n$. Claeskens et al. (2009) showed that $\|H^{-1}\|_\infty = O(1)$, where $H = \frac{K_n}{n}X^TX + \lambda D_m^T D_m$. Thus, $\hat{b}$ is stochastically bounded, so is $\hat{f}^{[p]}$. Let $\bar{b}$ solve $(X^TX + \lambda^* D_m^T D_m)\bar{b} = X^T f$ and denote $\bar{f}(x) = \sum_{k=1}^{K_n+p} \bar{b}_k B_k^{(p)}(x)$. We have

$$\|\hat{f}^{[p]} - \bar{f}\| \le \|\hat{b} - \bar{b}\|_\infty \le \|H^{-1}\|_\infty \| \bar{y} - \mathbb{E}[\bar{y}] \|_\infty = O_p\Big(\sqrt{\frac{K_n}{n}}\sqrt{2\log K_n}\Big). \tag{32}$$

It is shown that $\|\bar{f} - f\| = O(\alpha)$ if $p = m$. The development of this result is a special case of Theorem 4.1 in Section 4. Thus,

$$\begin{aligned}
& |\hat{F}_1(x) - G_1(x) + G_1(\kappa_{k_x+p}) - \check{F}_1(\kappa_{k_x+p})| \\
\le\; & |\hat{F}_1(x) - \check{F}_1(x)| + |(G_1(\kappa_{k_x+p}) - G_1(x)) - (\Phi_1(\kappa_{k_x+p}) - \Phi_1(x))| \\
& + |(\Phi_1(\kappa_{k_x+p}) - \Phi_1(x)) - (\bar{F}_1(\kappa_{k_x+p}) - \bar{F}_1(x))| + |(\bar{F}_1(\kappa_{k_x+p}) - \bar{F}_1(x)) - (\check{F}_1(\kappa_{k_x+p}) - \check{F}_1(x))| \\
\le\; & \frac{2}{n}\|\hat{f}\| + O_p\Big(\frac{1}{\sqrt{nK_n}}\Big) + \frac{pM_n}{n}\|\bar{f} - f\| + \frac{pM_n}{n}\|\hat{f} - \bar{f}\| \\
=\; & O_p\Big(\frac{1}{n}\Big) + O_p\Big((\frac{\log K_n}{nK_n})^{1/2}\Big) + O_p\Big(\frac{\alpha}{K_n}\Big).
\end{aligned}$$

A similar rate can be obtained for $|R_{y,k_x+1} - R_{f,k_x+1}|$. Given the admissible ranges of K_n and α in next Corollary 4.2, $O_p((\log K_n/nK_n)^{1/2})$ is the dominating term. Hence, the lemma follows. □

Theorem 4.1. *If the true regression function is $2m$th order continuously differentiable with bounded $2m$th derivative, then the P-spline estimator $\hat{f}^{[m]}$ can be written as*

$$\begin{aligned}
\hat{f}^{[m]}(t) &= f(t) + (-1)^{m-1}\alpha f^{(2m)}(t) + o(\alpha) + \frac{1}{n}\sum_{i=1}^n K(t,t_i)\epsilon_i \\
&\quad + O_p\Big(\sqrt{\frac{\log K_n}{nK_n}}\Big)\beta^m + e^{-\beta t(1-t)}O_p(\beta^m),
\end{aligned} \tag{33}$$

uniformly in α and in $t \in (0,1)$.

Proof. Taking the mth derivative of $\int_0^1 K(s,t)G_m(s)ds$, we obtain

$$\begin{aligned}\int_0^1 \frac{\partial^m K(t,s)}{\partial t^m} G_m(s)ds &= \int_0^1 \frac{\partial^m K(t,s)}{\partial t^m}\Phi_m(s)ds + \int_0^1 \frac{\partial^m K(t,s)}{\partial t^m}\Big[G_m(s) - \tilde{\Phi}_m(s)\Big]ds \\ &\quad + \int_0^1 \frac{\partial^m K(t,s)}{\partial t^m}\Big[\tilde{\Phi}_m(s) - \Phi_m(s)\Big]ds.\end{aligned}$$

It is easy to show that

$$\int_0^1 \frac{\partial K(t,s)}{\partial t}\Phi_m(s)ds = -\int_0^1 \frac{\partial K(t,s)}{\partial s}\Phi_m(s)ds = -\Phi_m(1)K(t,1) + \int_0^1 K(t,s)\Phi_{m-1}(s)ds.$$

Therefore,

$$\int_0^1 \frac{\partial^m K(t,s)}{\partial t^m}\Phi_m(s)ds = -\sum_{j=0}^{m-1}\Phi_{m-j}(1)\frac{\partial^{m-j}}{\partial t^{(m-j)}}K(t,1) + \int_0^1 K(t,s)f(s)ds.$$

By Equation (6.4) in Theorem 2.2 of Nychka (1995), we have

$$\int_0^1 K(t,s)f(s)ds = f(t) + (-1)^{m-1}\alpha f^{(2m)}(t) + o(\alpha).$$

Similarly,

$$\begin{aligned}\int_0^1 \frac{\partial^m K(t,s)}{\partial t^m}\Big[G_m(s) - \tilde{\Phi}_m(s)\Big]ds &= -\sum_{j=0}^{m-1}\Big[G_{m-j}(1) - \tilde{\Phi}_{m-j}(1)\Big]\frac{\partial^{m-j}}{\partial t^{(m-j)}}K(t,1) \\ &\quad + \int_0^1 K(t,s)\Big[dG_1(s) - d\tilde{\Phi}_1(s)\Big], \\ &= O_p\Big(\beta^m e^{\frac{-\beta(1-t)}{\sqrt{n}}}\Big) + \frac{1}{n}\sum_{i=1}^n K(t,t_i)\epsilon_i.\end{aligned}$$

Moreover,

$$\Big|\int_0^1 \frac{\partial^m K(t,s)}{\partial t^m}\Big[\tilde{\Phi}_m(s) - \Phi_m(s)\Big]ds\Big| \leq \|\tilde{\Phi}_m - \Phi_m\| \ \Big|\int_0^1 \frac{\partial^m K(t,s)}{\partial t^m}ds\Big|,$$

which is of order $O(1/n)\beta^m$. Finally, in light of Lemma 4.1, $\|\frac{d^m}{dt^m}\int_0^1 K(s,t)\tilde{R}(s)ds\|$ is of order $(\log K_n/nK_n)^{1/2}\beta^m$. It is easy to verify that the mth derivative of $J(t)$ is of order $e^{-\beta t(1-t)}\beta^m$. This completes the detail of the representation. □

Remark 4.1. Theorem 4.1 indicates that the P-spline estimator is approximately a kernel regression estimator. The equivalent kernel is $K(t,s)$ given in Section 3, and α plays a role similar to the bandwidth h. The asymptotic mean has the bias $(-1)^{m-1}\alpha f^{(2m)}(x)$, which can be negligible if α is reasonably small. On the other hand, α can not be arbitrarily small as that will inflate the random component. The admissible range for α given in Corollary 4.1 is a compromise between these two.

Corollary 4.1. *Let α satisfy $\alpha n^{2m/(4m+1)} \to 0$ and $\alpha^{-(2m-1)/2m} \log K_n/K_n \to 0$. Suppose also that the true regression function f is $2m$th order continuously differentiable with bounded $2m$th derivative. Then for $t \in (0,1)$,*

$$\sqrt{\frac{n}{\beta}}\ [\hat{f}^{[m]}(t) - f(t)] \to^d N\big(0, \sigma_K^2(t)\big), \tag{34}$$

where $\frac{1}{\beta}\int_0^1 K^2(t,s)ds \to \sigma_K^2(t)$ as $n \to \infty$. However, if $\alpha = c^{2m} n^{-\frac{2m}{4m+1}}$ for $c > 0$, and let $K_n \sim n^\gamma$ with $\gamma > (2m-1)/(4m+1)$, then

$$n^{2m/(4m+1)}\ [\hat{f}^{[m]}(t) - f(t)] \to^d N\Big((-1)^{m-1} c^{2m} f^{(2m)}(t),\ \ \frac{\sigma_K^2(t)}{c}\ \Big). \tag{35}$$

Proof. Let $\Pi(t) = \frac{1}{n}\sum_{i=1}^n K(t,t_i)\epsilon_i$. For any fixed t, the Lindeberg-Levy central limit theorem gives

$$\sqrt{\frac{n}{\beta}}\ \Pi(t) \to N(0, \sigma_K^2(t))$$

in distribution, where $\frac{1}{\beta}\int_0^1 K^2(t,s)ds \to \sigma_K^2(t)$ as $n \to \infty$. If α satisfies $\alpha n^{2m/(4m+1)} \to 0$ and $\alpha^{-(2m-1)/2m} \log K_n/K_n \to 0$, it is easy to see that the remainder terms in (33) are $o_p(1)$. If $\alpha = c^{2m} n^{-\frac{2m}{4m+1}}$ for $c > 0$, and $K_n \sim n^\gamma$ with $\gamma > (2m-1)/(4m+1)$, we have $\sqrt{n/\beta}\alpha = c^{2m+1/2}$ and $\sqrt{n/\beta}\sqrt{\frac{\log K_n}{nK_n}}\beta^m \to 0$. The theorem follows. □

Remark 4.2. The asymptotic results in Corollary 4.1 provide theoretical justification of the observation that the number of knots is not important, as long as it is above some minimal level (Ruppert, 2002). It is easy to find that the mean squared error of the P-spline estimator is of order $n^{-4m/4m+1}$, which achieves the optimal rate of convergence given in Stone (1982).

In the following, we study the asymptotic property of $\hat{f}^{[p]}(t) = \sum_{k=1}^{K_n+p} \hat{b}_k B_k^{[p]}(t)$ when $p \neq m$. We first define a piecewise mth degree polynomial $\tilde{f}^{[m]}$, where $\hat{f}^{[p]}$ and $\tilde{f}^{[m]}$ share the same set of spline coefficients. In particular, define $\tilde{f}^{[m]}(t) = \sum_{k=1}^{K_n+m} \hat{b}_k B_k^{[m]}(t)$ if $p > m$, or $\tilde{f}^{[m]}(t) = \sum_{k=1}^{K_n+p} \hat{b}_k B_k^{[m]}(t)$ if $p < m$. Note that, if $p < m$, $\tilde{f}^{[m]}$ is defined on $[0, 1 - \frac{m-p}{K_n}]$. Following the similar discussion as above, we can establish the asymptotic distribution for $\tilde{f}^{[m]}$ as in (34) and (35), respectively, under different admissible ranges of α and K_n.

Lemma 4.2. *For any $t \in (0,1)$, let $d = \lfloor K_n t \rfloor + 1$. Let $\hat{\gamma}(t) = \hat{f}^{[p]}(t) - \tilde{f}^{[m]}(t)$. Then, if $p > m$,*

$$\hat{\gamma}(t) = \sum_{q=m+1}^{p} \sum_{i=d+1}^{d+q} \Big(\frac{K_n}{q}(t - \kappa_{i-q})\Big) B_i^{[q-1]}(t) \sum_{l=1}^{p} a_{i+1-d,l} K_n^{-l} \frac{d^l}{dt^l} \hat{f}^{[p]}(t), \tag{36}$$

and if $p < m$,

$$\hat{\gamma}(t) = -\sum_{q=p+1}^{m} \sum_{i=d+1}^{d+m} \Big(\frac{K_n}{q}(t - \kappa_{i-q})\Big) B_i^{[q-1]}(t) \sum_{l=1}^{m} b_{i+1-d,l} K_n^{-l} \frac{d^l}{dt^l} \tilde{f}^{[m]}(t), \tag{37}$$

where the coefficients $\{a_{ij}\}$ and $\{b_{ij}\}$ are constants.

Proof. The B-spline basis functions have the recurrence relationship such that

$$B_j^{[p]}(t) = \frac{K_n}{p}(t - \kappa_{j-p-1})B_{j-1}^{[p-1]}(t) + \frac{K_n}{p}(\kappa_j - t)B_j^{[p-1]}(t).$$

Let $f^{[p-1]}(t) = \sum_{k=1}^{K_n+p-1} b_k B_k^{[p-1]}(t)$ with the same first $(K_n + p - 1)$ coefficients of $f^{[p]}$. For $x \in (\kappa_d, \kappa_{d+1})$, the difference between $f^{[p]}(t)$ and $f^{[p-1]}(t)$ is given by

$$\begin{aligned} f^{[p]}(t) - f^{[p-1]}(t) &= \sum_{i=d+1}^{d+p} \Big[b_{i+1}\frac{K_n}{p}(t - \kappa_{i-p}) + b_i\big(\frac{K_n}{p}(\kappa_i - t) - 1\big)\Big] B_i^{[p-1]}(t) \\ &= \sum_{i=d+1}^{d+p} (b_{i+1} - b_i)\Big(\frac{K_n}{p}(t - \kappa_{i-p})\Big) B_i^{[p-1]}(t). \end{aligned} \tag{38}$$

From (38), if $p > m$,

$$\hat{f}^{[p]}(t) = \tilde{f}^{[m]}(t) + \sum_{q=m+1}^{p} \sum_{i=d+1}^{d+q} \Delta b_{i+1}\Big(\frac{K_n}{q}(t - \kappa_{i-q})\Big) B_i^{[q-1]}(t).$$

From (2), we have $\Delta^l b_k = c_l^T(\Delta b_{k-l+1}, \Delta b_{k-l+2}, \ldots, \Delta b_k)$, where

$$c_l = \Big[\ (-1)^{l-1}\binom{l-1}{0}, (-1)^{l-2}\binom{l-1}{1}, \ldots, (-1)^0\binom{l-1}{l-1}\ \Big]^T.$$

Combining this with (7), it is easy to show that there exists $C_d \in \mathbb{R}^{p\times p}$ such that

$$\Big[\ \Delta b_{d+2}, \Delta b_{d+2}, \ldots, \Delta b_{d+p+1}\ \Big]^T = C_d\ \Big[\ K_n^{-1}\frac{d}{dt}f^{[p]}(t), \ldots, K_n^{-p}\frac{d^p}{dt^p}f^{[p]}(t)\ \Big]^T.$$

Hence, we can write

$$\Delta b_{d+k} = \sum_{l=1}^{p} a_{kl}K_n^{-l}\frac{d^l}{dx^l}f^{[p]}(t),\ k = 2, \ldots, p+1, \tag{39}$$

which gives (36). (37) can be established similarly. Thus the lemma follows. □

Corollary 4.2. *Suppose that f is $2m$th order continuously differentiable with bounded $2m$th derivative on $[0,1]$. Let α satisfy $\alpha n^{2m/(4m+1)} \to 0$ and $\alpha^{-(2m-1)/2m}\log K_n/K_n \to 0$. Then, for $t \in (0,1)$,*

$$\sqrt{\frac{n}{\beta}}\ [\hat{f}^{[p]}(t) - f(t) - \hat{\gamma}(t)] \to^d N\big(0, \sigma_K^2(t)\big), \tag{40}$$

where $\gamma(t)$ is given by (36) if $p < m$ or (37) if $p > m$. However, if $\alpha = c^{2m}n^{-\frac{2m}{4m+1}}$ for $c > 0$, and let $K_n \sim n^\gamma$ with $\gamma > (2m-1)/(4m+1)$, then

$$n^{2m/(4m+1)}\ [\hat{f}^{[p]}(t) - f(t) - \hat{\gamma}(t)] \to^d N\Big((-1)^{m-1}c^{2m}f^{(2m)}(t),\ \ \frac{\sigma_K^2(t)}{c}\ \Big). \tag{41}$$

Remark 4.3. When p is not equal to m, the asymptotic bias has an additional term $\hat{\gamma}(t)$, which is of order $O_p(1/K_n)$. When K_n grows sufficiently fast with respect to n, this term is asymptotically negligible.

5 The equivalent kernels near boundary

The approximation of the equivalent kernel $K(t,s)$ deteriorates when t is near the boundary points of the design set. In this section, we derive an explicit formula for the equivalent kernel when t is close to the boundary. We discuss the case when t is close to 0 only; the case when t is close to 1 follows from the similar argument and thus is omitted.

Consider an even m first. It follows from the closed-form expressions (23) and (24) for $\hat{F}_m(t)$ that for $t \in [0,1]$ sufficiently small, the m-th derivative of $\sum_{k=0}^{\frac{m}{2}-1} e^{-\beta\mu_k(1-t)}\big[a_k^+ \cos(\beta\omega_k t) + b_k^+ \sin(\beta\omega_k t)\big]$ is of order $O_p(\beta^m e^{-\beta})$. Hence, we only consider

$$\widetilde{F}(t) \equiv F_0(t) + \sum_{k=0}^{\frac{m}{2}-1} e^{-\beta\mu_k t}\big[a_k \cos(\beta\omega_k t) + b_k \sin(\beta\omega_k t)\big]. \tag{42}$$

In the subsequent, we shall express the coefficients a_k, b_k in terms of $F_0(0)$ and its derivatives. This will eventually lead to an explicit expression for the kernel.

In view of (15), we have

$$F_0^{(j)}(0) = \int_0^1 \frac{\partial L^{(j)}(|s-t|)}{\partial\, t^j}\Big|_{t=0} G(s)ds = (-1)^j \int_0^1 \frac{\partial L^{(j)}(s)}{\partial\, s^j} G(s)ds, \quad \forall\ j = 1, \cdots, 2m-1. \tag{43}$$

Moreover, it follows from Section 3.1 that $L(s) = \beta \sum_{k=0}^{\frac{m}{2}-1} p_k(s) = \beta \sum_{k=0}^{\frac{m}{2}-1} \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} p_k(s) \\ q_k(s) \end{bmatrix}$, where

$$\begin{bmatrix} p_k(s) \\ q_k(s) \end{bmatrix} = e^{-\beta\mu_k s} S(\omega_k \beta s) \begin{bmatrix} c_k \\ d_k \end{bmatrix},$$

and $S(\cdot) \in \mathrm{SO}(2)$ is given by $S(\cdot) = \begin{bmatrix} \cos(\cdot) & \sin(\cdot) \\ -\sin(\cdot) & \cos(\cdot) \end{bmatrix}$. In light of (17), we have

$$\begin{bmatrix} p_k^{(j)}(s) \\ q_k^{(j)}(s) \end{bmatrix} = (\beta A_k)^j \begin{bmatrix} p_k(s) \\ q_k(s) \end{bmatrix} = (\beta A_k)^j e^{-\beta\mu_k s} S(\omega_k \beta s) \begin{bmatrix} c_k \\ d_k \end{bmatrix}.$$

As a result, we obtain

$$\frac{\partial L^{(j)}(s)}{\partial\, s^j} = \beta \sum_{k=0}^{\frac{m}{2}-1} \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} p_k^{(j)}(s) \\ q_k^{(j)}(s) \end{bmatrix} = \beta^{j+1} \sum_{k=0}^{\frac{m}{2}-1} (A_k)^j_{1\bullet} e^{-\beta\mu_k s} S(\omega_k \beta s) \begin{bmatrix} c_k \\ d_k \end{bmatrix},$$

where $(A_k)^j_{1\bullet}$ denotes the first row of the j-th power of A_k given in (17). This, along with (43), yields

$$\begin{aligned} \frac{F_0^{(j)}(0)}{\beta^j} &= (-1)^j \int_0^1 \Big(\sum_{k=0}^{\frac{m}{2}-1} (A_k)^j_{1\bullet} \beta e^{-\beta\mu_k s} S(\omega_k \beta s) \begin{bmatrix} c_k \\ d_k \end{bmatrix} \Big) G(s)ds \\ &= \sum_{k=0}^{\frac{m}{2}-1} (-A_k)^j_{1\bullet} \int_0^1 \Big(\beta e^{-\beta\mu_k s} S(\omega_k \beta s) \begin{bmatrix} c_k \\ d_k \end{bmatrix} \Big) G(s)ds. \end{aligned}$$

For notational simplicity, let $\widehat{B}_{11}^{e}$ denote the inverse of the matrix B_{11}^{e} defined in (26) and let

$$\mathbf{p}_j = \sum_{k=0}^{\frac{m}{2}-1} (-A_k)_{1\bullet}^{j} \int_0^1 \Big(\beta e^{-\beta\mu_k s} S(\omega_k \beta s) \begin{bmatrix} c_k \\ d_k \end{bmatrix} \Big) G(s) ds, \quad j = 0, \cdots, m-1,$$

and $\mathbf{p} = [\mathbf{p}_0, \mathbf{p}_1, \cdots, \mathbf{p}_{m-1}]^T \in \mathbb{R}^m$. Therefore, it follows from the development in Section 3.4 that

$$\begin{bmatrix} a_0 \\ b_0 \\ \vdots \\ a_{\frac{m}{2}-1} \\ b_{\frac{m}{2}-1} \end{bmatrix} = -(\widehat{B}_{11}^{e} + O_p(e^{-\beta}))\, \mathbf{p} \;=\; -\widehat{B}_{11}^{e}\, \mathbf{p} + O_p(e^{-\beta}). \tag{44}$$

Returning to $\widetilde{F}^{(m)}(t)$ and using (44), we have, for $\beta \to \infty$,

$$\begin{aligned} \widetilde{F}^{(m)}(t) \;&=\; F_0^{(m)}(t) + \beta^m \sum_{\ell=0}^{\frac{m}{2}-1} (A_\ell)_{1\bullet}^{m} e^{-\beta\mu_\ell t} S(\omega_\ell \beta\, t) \begin{bmatrix} a_\ell \\ b_\ell \end{bmatrix} \\ &=\; \int_0^1 \frac{\partial L^{(m)}(|t-s|)}{\partial\, s^m} G(s) ds \;+\; \beta^m\, \mathbf{q}^T(t) \Big(-\widehat{B}_{11}^{e} \Big) \mathbf{p}, \end{aligned} \tag{45}$$

where $\mathbf{q}(t) \equiv \big[\mathbf{q}_0(t), \mathbf{q}_1(t), \cdots, \mathbf{q}_{\frac{m}{2}-1}(t)\big]^T \in \mathbb{R}^m$ with $\mathbf{q}_\ell(t) \equiv (A_\ell)_{1\bullet}^{m} e^{-\beta\mu_\ell t} S(\omega_\ell \beta t) \in \mathbb{R}^{1\times 2}$ for $\ell = 0, 1, \cdots, \frac{m}{2}-1$.

To find the kernel in this case, particularly the kernel for the second term, recall

$$\begin{bmatrix} p_k(s) \\ q_k(s) \end{bmatrix} = e^{-\beta\mu_k s} S(\omega_k \beta s) \begin{bmatrix} c_k \\ d_k \end{bmatrix}.$$

Therefore, the second term in (45) becomes

$$\beta^m\, \mathbf{q}^T(t) \Big(-\widehat{B}_{11}^{e} \Big) \mathbf{p} \;=\; \int_0^1 W(t,s) G(s) ds,$$

where

$$W(t,s) = \beta^m \mathbf{q}^T(t) \Big(-\widehat{B}_{11}^{e} \Big) \begin{bmatrix} \nu_0(s) \\ \nu_1(s) \\ \vdots \\ \nu_{m-1}(s) \end{bmatrix}$$

and $\nu_j(s) = \sum_{k=0}^{\frac{m}{2}-1} (-A_k)_{1\bullet}^{j} \Big(\beta \begin{bmatrix} p_k(s) \\ q_k(s) \end{bmatrix} \Big)$, $j = 0, \ldots, m-1$.

Denote by $\overline{p}_k^{(r)}(s)$ and $\overline{q}_k^{(r)}(s)$ the r-th order integrals of $p_k(s)$ and $q_k(s)$ respectively, namely,

$$\overline{p}_k^{(r)}(s) \equiv \underbrace{\int \cdots \int}_{r-\text{copies}} p_k\, d\tau_1 \cdots d\tau_r, \qquad \overline{q}_k^{(r)}(s) \equiv \underbrace{\int \cdots \int}_{r-\text{copies}} q_k\, d\tau_1 \cdots d\tau_r.$$

In light of (17), it is easy to verify that

$$\begin{bmatrix} \overline{p}_k^{(r)}(s) \\ \overline{q}_k^{(r)}(s) \end{bmatrix} = (\beta A_k)^{-r} \begin{bmatrix} p_k(s) \\ q_k(s) \end{bmatrix} = e^{-\beta\mu_k s}(\beta A_k)^{-r} S(\omega_k \beta s) \begin{bmatrix} c_k \\ d_k \end{bmatrix}.$$

Using this and $\dfrac{\partial^m \overline{p}_k^{(m)}(s)}{\partial s^m} = p_k(s)$, $\dfrac{\partial^m \overline{q}_k^{(m)}(s)}{\partial s^m} = q_k(s)$, we have

$$\begin{aligned} \nu_j(s) &= \sum_{k=0}^{\frac{m}{2}-1} (-A_k)^j_{1\bullet} \Big(\beta \begin{bmatrix} \frac{\partial^m \overline{p}_k^{(m)}(s)}{\partial s^m} \\ \frac{\partial^m \overline{q}_k^{(m)}(s)}{\partial s^m} \end{bmatrix} \Big) = \frac{\partial^m}{\partial s^m} \sum_{k=0}^{\frac{m}{2}-1} (-A_k)^j_{1\bullet} \Big(\beta \begin{bmatrix} \overline{p}_k^{(m)}(s) \\ \overline{q}_k^{(m)}(s) \end{bmatrix} \Big) \\ &= \frac{\partial^m}{\partial s^m} \sum_{k=0}^{\frac{m}{2}-1} (-A_k)^j_{1\bullet} \Big(\beta e^{-\beta\mu_k s} (\beta A_k)^{-m} S(\omega_k \beta s) \begin{bmatrix} c_k \\ d_k \end{bmatrix} \Big). \end{aligned}$$

Therefore,

$$W(t,s) \equiv \frac{\partial^m}{\partial s^m} \Big(\mathbf{q}^T(t) \big(-\widehat{B}^e_{11} \big) \, \mathbf{r}(s) \Big), \tag{46}$$

where $\mathbf{r}(s) = [\mathbf{r}_0(s), \mathbf{r}_1(s), \ldots, \mathbf{r}_{m-1}(s)]^T \in \mathbb{R}^m$, and

$$\mathbf{r}_j(s) = \sum_{k=0}^{\frac{m}{2}-1} (-A_k)^{-(m-j)}_{1\bullet} \beta e^{-\beta\mu_k s} S(\omega_k \beta s) \begin{bmatrix} c_k \\ d_k \end{bmatrix}, \qquad j = 0, \ldots, m-1.$$

Here the coefficients c_k, d_k satisfy the linear equation (18). Finally, we obtain the equivalent kernel for $t \geq 0$ sufficiently small (when $\beta \to \infty$) as

$$K_b(t,s) = L(|t-s|) + \mathbf{q}^T(t) \big(-\widehat{B}^e_{11} \big) \, \mathbf{r}(s). \tag{47}$$

Example 5.1. As an illustration, we derive the closed-form expression of the kernel near the boundary $t = 0$ for $m = 2$ and compare it with the boundary kernel established by Silverman (1984) for the smoothing splines. Since $c_0 = d_0 = \frac{1}{2\sqrt{2}}$,

$$A_0 = \begin{bmatrix} \cos(\frac{3\pi}{4}) & \sin(\frac{3\pi}{4}) \\ -\sin(\frac{3\pi}{4}) & \cos(\frac{3\pi}{4}) \end{bmatrix}, \quad S(\omega_0 \beta t) = \begin{bmatrix} \cos(\frac{\beta t}{\sqrt{2}}) & \sin(\frac{\beta t}{\sqrt{2}}) \\ -\sin(\frac{\beta t}{\sqrt{2}}) & \cos(\frac{\beta t}{\sqrt{2}}) \end{bmatrix}, \quad \widehat{B}^e_{11} = \begin{bmatrix} 1 & 0 \\ 1 & \sqrt{2} \end{bmatrix},$$

we have

$$\mathbf{q}(t) = e^{-\frac{\beta t}{\sqrt{2}}} \begin{bmatrix} \sin(\frac{\beta t}{\sqrt{2}}) \\ -\cos(\frac{\beta t}{\sqrt{2}}) \end{bmatrix}, \qquad \mathbf{r}(t) = \frac{\beta}{2\sqrt{2}} e^{-\frac{\beta s}{\sqrt{2}}} \begin{bmatrix} -\sin(\frac{\beta s}{\sqrt{2}}) + \cos(\frac{\beta s}{\sqrt{2}}) \\ \sqrt{2}\cos(\frac{\beta s}{\sqrt{2}}) \end{bmatrix}.$$

Hence, the equivalent kernel near the boundary $t = 0$ is

$$\begin{aligned} K_b(t,s) &= L(|t-s|) + \mathbf{q}^T(t) \Big(-\widehat{B}^e_{11} \Big) \, \mathbf{r}(s) \qquad (48) \\ &= L(|t-s|) + \frac{\beta}{2\sqrt{2}} e^{-\frac{\beta}{\sqrt{2}}(t+s)} \Big[\cos\big(\frac{\beta}{\sqrt{2}}(t-s)\big) + 2\cos\big(\frac{\beta}{\sqrt{2}}t\big)\cos\big(\frac{\beta}{\sqrt{2}}s\big) - \sin\big(\frac{\beta}{\sqrt{2}}(t+s)\big) \Big]. \end{aligned}$$

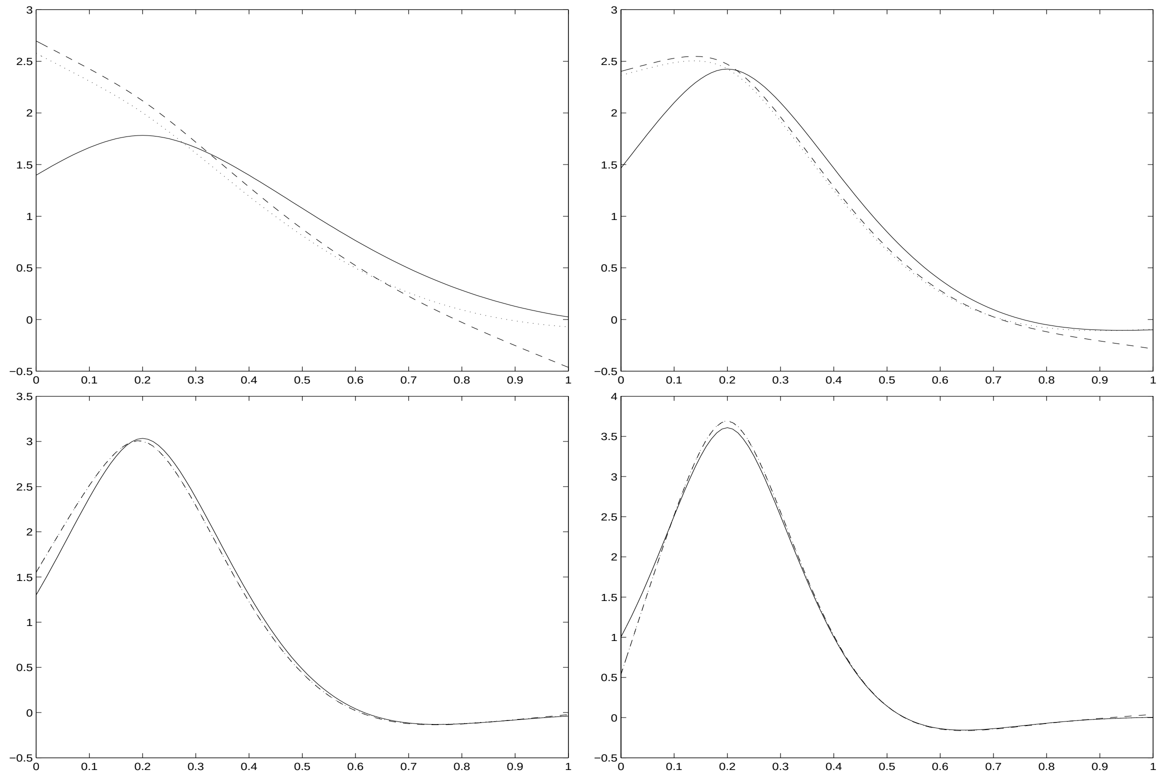

Figure 2: The non-boundary kernel (solid), the finite-sample kernel (dashed) and asymptotic boundary kernel (dotted) for $m = 2$ and for (a) $\beta = 4$, (b) $\beta = 6$, (c) $\beta = 8$, (d) $\beta = 10$. The kernels are for estimation at $x = 0.2$.

When $t=0$, since $L(|t-s|)=\frac{\beta}{2\sqrt{2}}\,e^{-\frac{\beta}{\sqrt{2}}s}\big[\cos(\frac{\beta}{\sqrt{2}}s)+\sin(\frac{\beta}{\sqrt{2}}s)\big]$, the boundary kernel becomes $\sqrt{2}\,e^{-\frac{\beta}{\sqrt{2}}s}\cos(\frac{\beta}{\sqrt{2}}s),\ s\in[0,1]$. It is interesting to notice that the boundary kernel in (48) agrees with that obtained by Silverman (1984). Figure 2 displays the non-boundary kernel, boundary kernel, and the finite sample kernel when we estimate $x=0.2$ with different choices of β, where the finite sample kernel is obtained by incorporating the terms containing $e^{-\beta\mu_k(1-t)}$ ignored in (42). Indeed, this kernel is given by

$$
\begin{aligned}
&L(|t-s|)\ +\ \tfrac{\beta}{2\sqrt{2}}\,e^{-\frac{\beta}{\sqrt{2}}(t+s)}\Big[\cos\big(\tfrac{\beta}{\sqrt{2}}(t-s)\big)+2\cos\big(\tfrac{\beta}{\sqrt{2}}t\big)\cos\big(\tfrac{\beta}{\sqrt{2}}s\big)-\sin\big(\tfrac{\beta}{\sqrt{2}}(t+s)\big)\Big]\\
&+\tfrac{\beta}{2}\,e^{-\frac{\beta}{\sqrt{2}}(2-t-s)}\Big\{\cos\big(\tfrac{\beta(1-t)}{\sqrt{2}}+\tfrac{\pi}{4}\big)\Big[\cos(\tfrac{\beta(1-s)}{\sqrt{2}})-\sin(\tfrac{\beta(1-s)}{\sqrt{2}})\Big]+\sqrt{2}\cos(\tfrac{\beta(1-t)}{\sqrt{2}})\cos(\tfrac{\beta(1-s)}{\sqrt{2}})\Big\}.
\end{aligned}
$$

There are a good agreement between the finite-sample and asymptotic kernels when $\beta=6$ and an excellent agreement when $\beta=10$.

The development of the boundary kernel for an odd m is similar and we omit the details here. For notational simplicity, let $\widehat{B}_{11}^{o}$ denote the inverse of the matrix B_{11}^{o} defined in (30). we obtain the equivalent kernel for $t\geq 0$ sufficiently small (when $\beta\to\infty$) as

$$
K_b(t,s)=P(|t-s|)+\mathbf{q}^T(t)\Big(-\widehat{B}_{11}^{o}\Big)\mathbf{r}(s), \tag{49}
$$

where $\mathbf{q}(t)=\big[(-1)^m e^{-\beta t},\mathbf{q}_1(t),\ldots,\mathbf{q}_{\frac{m-1}{2}}(t)\big]^T\in\mathbb{R}^m$ and $\mathbf{r}(s)=[\mathbf{r}_0(s),\mathbf{r}_1(s),\ldots,\mathbf{r}_{m-1}(s)]^T\in\mathbb{R}^m$, and

$$
\begin{aligned}
\mathbf{q}_\ell(t) &= (A_\ell)^m_{1\bullet}e^{-\beta\mu_\ell t}S(\omega_\ell\beta t)\in\mathbb{R}^{1\times 2},\qquad \ell=1,\ldots,\frac{m-1}{2},\\
\mathbf{r}_j(s) &= (-1)^m c_0\beta e^{-\beta s}+\sum_{k=1}^{\frac{m-1}{2}}(-A_k)^{-(m-j)}_{1\bullet}\beta e^{-\beta\mu_k s}S(\omega_k\beta s)\begin{bmatrix}c_k\\ d_k\end{bmatrix},\quad j=0,\ldots,m-1,
\end{aligned}
$$

where the coefficients c_k,d_k satisfy the linear equation (21).

6 Extensions to unequally spaced data and multivariate smoothing

We have so far focused on the equally spaced design case and equally spaced knots. When the design is not equally spaced, one can use the ideas of Stute (1984) and Li and Ruppert (2008). In specific, assume that x_i's are in (a,b). Find a smoothing monotone function Υ such that $\Upsilon(x_i)=i/n$ from (a,b) to $(0,1)$. We use the P-spline smoothing to fit $(i/n,y_i)$, and thus the regression function is give by $f\circ\Upsilon^{-1}$. We place knots at sample quantiles so that there are equal numbers of data points between consecutive knots.

The univariate P-splines can be naturally extended to multivariate P-splines (Marx and Eilers, 2005). The asymptotic properties can be studied along the same line. Consider the problem of estimating the ν dimensional function $f(t_1,\ldots,t_\nu)$ from noisy observations $y_i=f(t_{1i},\ldots,t_{\nu i})+\epsilon_i$, $i=1,\ldots,n$. The P-spline model approximates f by

$$
f(t_1,\ldots,t_\nu)=\sum_{k_1=1}^{K_{1n}+p_1}\cdots\sum_{k_\nu=1}^{K_{\nu n}+p_\nu}b_{k_1,\ldots,k_\nu}B_{k_1}^{[p_1]}(t_1)\cdots B_{k_\nu}^{[p_\nu]}(t_\nu).
$$

The spline coefficient $\hat{b}$ subject to the difference penalty are chosen to minimize

$$\sum_{i=1}^{n}\big[y_i - \sum_{k_1=1}^{K_{1n}+p_1}\cdots\sum_{k_d=1}^{K_{dn}+p_d} b_{k_1,\dots,k_d}B_{k_1}^{[p_1]}(t_{1i})\cdots B_{k_d}^{[p_d]}(t_{di})\big]^2 + \lambda^*\sum_{k_1=m_1+1}^{K_{1n}+p_1}\cdots\sum_{k_d=m_d+1}^{K_{dn}+p_d}\big[\Delta^{m_1,m_2,\dots,m_d}b_{k_1,k_2,\dots,k_d}\big]^2,$$

where the difference operator for d dimensional case is defined as follows:

$$\begin{aligned}
\Delta^{0,\dots,0}b_{k_1,\dots,k_\nu} &= b_{k_1,\dots,k_\nu}, \quad k_1=1,\dots,K_{1n}+p_1,\dots,k_\nu=1,\dots,K_{\nu n}+p_\nu,\\
\Delta^{m_1,m_2,\dots,m_\nu}b_{k_1,k_2,\dots,k_\nu} &= \Delta^{m_1-1,m_2,\dots,m_\nu}b_{k_1,k_2,\dots,k_\nu} - \Delta^{m_1-1,m_2,\dots,m_\nu}b_{k_1-1,k_2,\dots,k_\nu}\\
&= \Delta^{m_1,m_2-1,\dots,m_\nu}b_{k_1,k_2,\dots,k_\nu} - \Delta^{m_1,m_2-1,\dots,m_\nu}b_{k_1,k_2-1,\dots,k_\nu}\\
&= \cdots\\
&= \Delta^{m_1,m_2,\dots,m_\nu-1}b_{k_1,k_2,\dots,k_\nu} - \Delta^{m_1,m_2,\dots,m_\nu-1}b_{k_1-1,k_2,\dots,k_\nu-1}.
\end{aligned}$$

For example, consider a two dimensional difference operator when $k_1=1$ and $k_2=2$:

$$\begin{aligned}
\Delta^{1,2}b_{ks} &= \Delta^{0,2}b_{ks} - \Delta^{0,2}b_{k-1,s}\\
&= [b_{ks}-2b_{k,s-1}+b_{k,s-2}] - [b_{k-1,s}-2b_{k-1,s-1}+b_{k-1,s-2}].
\end{aligned}$$

Let X be the $n\times\{\Pi_{j=1}^{\nu}(K_{jn}+p_j)\}$ matrix with (i,j)th entry equal to $B_{k_1}^{[p_1]}(t_{1i})\cdots B_{k_\nu}^{[p_\nu]}(t_{di})$. Define D as the $\{\Pi_{j=1}^{\nu}(K_{jn}+p_j-m_j)\}\times\{\Pi_{j=1}^{\nu}(K_{jn}+p_j)\}$ differencing matrix satisfying

$$Db = \begin{pmatrix} \Delta^{m_1,\dots,m_\nu}b_{m_1+1,\dots,m_\nu+1}\\ \vdots \\ \Delta^{m_1,\dots,m_\nu}b_{K_{1n}+p_1,\dots,K_{\nu n}+p_\nu}\end{pmatrix}.$$

The optimality condition is given by

$$(X^TX+\lambda^*D^TD)\hat{b} = X^Ty. \tag{50}$$

Note that $D = D_{m_1}\otimes D_{m_2}\otimes\cdots\otimes D_{m_\nu}$ and $D^TD = D_{m_1}^TD_{m_1}\otimes D_{m_2}^TD_{m_2}\otimes\cdots\otimes D_{m_d}^TD_{m_d}$, where "$\otimes$" represents the Kronecker product. We may go though the same procedure as described in this paper. The multivariate P-spline smoothing is asymptotically equivalent to kernel smoothing and the equivalent kernel is the Green's function corresponding to the partial differential equation (PDE):

$$(-1)^{m_1+\cdots+m_d}\alpha\frac{\partial^{2m_1+\cdots+2m_d}}{\partial t_1^{2m_1}\cdots\partial t_d^{2m_d}}F(t_1,\dots,t_d) + F(t_1,\dots,t_d) = G(t_1,\dots,t_d), \tag{51}$$

subject to the boundary conditions:

$$\begin{aligned}
&\frac{\partial^{k_1+\cdots+k_d}}{\partial t_1^{k_1}\cdots\partial t_d^{k_d}}F(t_1,\dots,t_d) = 0, \text{ if any } t_i=0, \quad k_i=0,\dots,m_i-1,\\
&\frac{\partial^{k_1+\cdots+k_d}}{\partial t_1^{k_1}\cdots\partial t_d^{k_d}}F(t_1,\dots,t_d) = \frac{\partial^{k_1+\cdots+k_d}}{\partial t_1^{k_1}\cdots\partial t_d^{k_d}}G(t_1,\dots,t_d), \text{ if any } t_i=1, \quad k_i=0,\dots,m_i-1.
\end{aligned}$$

Further study of this issue is beyond the scope of this paper and shall be reported in a future publication.

References

[1] Claeskens, G., Krivobokova, T. and Opsomer, J. (2009). Asymptotic properties of penalized spline estimators. *Biometrika*, in print.

[2] De Boor, C. (2001) *A Pratical Guide to Splines*. Springer.

[3] Hall, P. and Opsomer, J.D. (2005). Theory for penalised spline regression. *Biometrika*, **92**, 105-118.

[4] Li, Y. and Ruppert, D. (2008). On the asymptotics of penalized splines. *Biometrika*, **95**, 415-436.

[5] Mammen, E. (1991). Estimating a smoothing regression function. *Annals of Statistics*, **19**, 724-740.

[6] Marx, B. and Eilers, P. (1996). Flexible smoothing with B-splines and penalties (with comments and rejoinder). *Statistical Science*, **11**, 89-121.

[7] Marx, B. and Eilers, P. (2005). Multidimensional penalized signal regression. *Technometrics*, **47** 13-22.

[8] Messer, K. (1991). A comparison of a spline estimate to its equivelent kernel estimate. *Annals of Statistics*, **19**, 817-829.

[9] Nychka, D. (1995). Splines as local smoothers. *Annals of Statistics*, **23**, 1175-1197.

[10] O'Sullivan, F. (1986). A statistical perspective on ill-posed inverse problems (with Discussion), *Statistical Science*, **1**, 505-527.

[11] Pal, J. and Woodroofe, M. (2007). Large sample properties of shape restricted regression estimators with smoothness adjuctments. *Statistica Sinica*, **17**, 1601-1616.

[12] Rice, J. and Rosenblatt, M. (1983). Smoothing splines: regression, derivatives and deconvolution. *Annals of Statistics*, **11**, 141-156.

[13] Ruppert, D. (2002). Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statisitcs*, **11**, 735-757.

[14] Ruppert, D. and Carroll, R. (2000). Spatially-adaptive penalities for spline fitting. *Australian & New Zealand Journal of Statistics*, **42**, 205-224.

[15] Ruppert, D., Wand, M.P., and Carroll, R.J. (2003). *Semiparametric Regression*. Cambridge: Cambridge University Press.

[16] Silverman, B.W. (1984). Spline smoothing: the equivalent variable kernel method. *Annals of Statistics*, **12**, 898-916.

[17] Shen, J. and Wang, X. (2009). Estimation of shape constrained functions in dynamical systems and its application to genetic networks. Submitted to 2010 American Control Conference.

[18] Stone, C.J. (1982). Optimal rate of convergence for nonparametric regression. *Annals of Statistics*, **10**, 1040-1053.

[19] Stute, W. (1984). Asymptotic normality of nearest neighbor regression function estimates. *Annals of Statistics*, **12**, 917-926.

[20] Wang, X. and Shen, J. (2009). On the asymptotics of monotone P-splines. Manuscript.